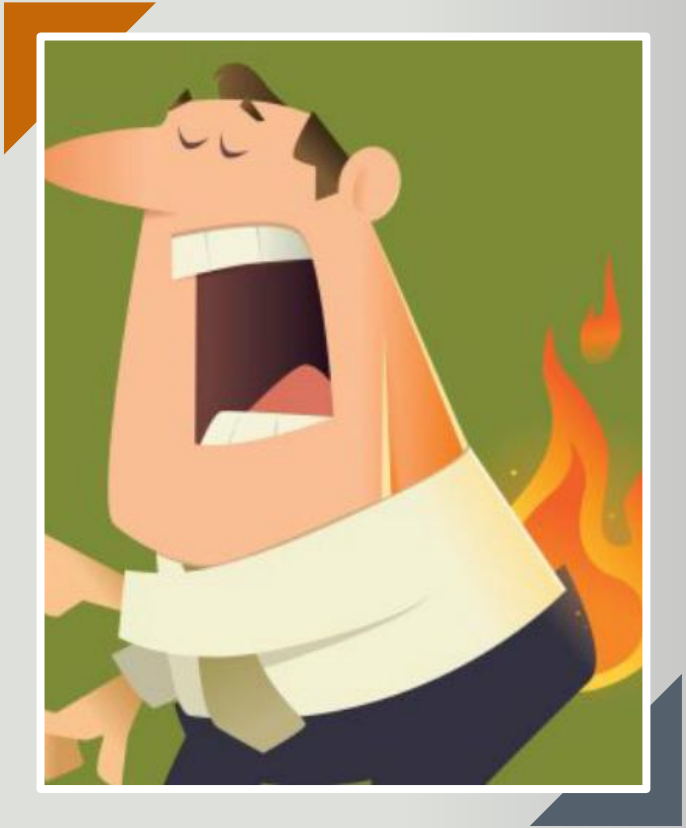

Liar Liar Pants on Fire

A New Benchmark Dataset for Fake News Detection

by

William Yang Wang

Implementation

- Preprocessing dataset
- Applying different models
- Comparing the results

Preprocessing

- Different models where applied to the dataset considering six target labels but many of these models didn’t got enough accuracy.
- The dataset was preprocessed to convert the 6 target labels to binary labels as:

```
'true': 'True',    'mostly-true': 'True',       'half-true': 'True',
'false': 'False',  'barely-true': 'False',     'pants-fire': 'False'
```

- Stopwords such as me, you, etc. are removed in the pipeline for each model.

Models Used

- **Logistic Regression**
- **SVM**
- **Random Forest**
- **Voting Classifier**
- **CNN**
- **LSTM**

Logistic Regression

- Multiclass
  Our Accuracy : 0.236
  Paper’s Accuracy:0.247

- Binary
  Accuracy : 0.611
  F1 score is: 0.666
  Precision score is: 0.644
  Recall score is: 0.689

SVM

- Multiclass
  Our Accuracy: 0.257
  Paper’s Accuracy: 0.255

- Binary
  Our Accuracy : 0.624
  F1 score is: 0.695
  Precision score is: 0.640
  Recall score is: 0.760

Random Forest

- Multiclass
  Our Accuracy : 0.212

- Binary
  Our Accuracy : 0.565
  F1 score is: 0.721
  Precision score is: 0.564
  Recall score is: 1.0

Voting Classifier

- Multiclass
  Our Accuracy : 0.246

- Binary
  Our Accuracy : 0.566
  F1 score is: 0.721
  Precision score is: 0.565
  Recall score is: 0.994

CNN

Our Accuracy: 0.235
Paper's Accuracy:0.260

- Filter size: (2, 3, 4)
- Number of filters:128
- Dropout probabilities: 0.8
- Batch size for stochastic gradient: 64

LSTM

Our Accuracy: 0.265
Paper’s Accuracy: 0.233